

Michael Beißwenger / Eva Gredel

# SPRACHKORPORA ALS DIGITALE BILDUNGSTECHNOLOGIEN: MODELLE FÜR DAS FORSCHENDE LERNEN IM DAF- UND DEUTSCHUNTERRICHT

## EINE ARBEITSGEMEINSCHAFT AM CENTER FOR ADVANCED INTERNET STUDIES NRW

Michael Beißwenger ist Professor für Germanistische Linguistik und Sprachdidaktik an der Universität Duisburg-Essen. Er ist Co-Leiter der Arbeitsgemeinschaft „Korpora als digitale Bildungstechnologien“ am CAIS NRW.

Eva Gredel ist Juniorprofessorin für Germanistische Linguistik mit dem Schwerpunkt „Digitale Kommunikation in Vermittlungskontexten“ an der Universität Duisburg-Essen. Sie ist Co-Leiterin der Arbeitsgemeinschaft „Korpora als digitale Bildungstechnologien“ am CAIS NRW.

### Einleitung

Wie können Korpora als digitale Bildungstechnologien in verschiedenen Vermittlungskontexten gewinnbringend genutzt werden? Dieser Frage geht die Arbeitsgemeinschaft (AG) „Sprachkorpora als digitale Bildungstechnologien“ nach, die im September 2022 am Center for Advanced Internet Studies (CAIS NRW) in Bochum gestartet ist. Als Institut für Digitalisierungsforschung bringt das CAIS Wissenschaftler:innen und Akteure aus der Praxis zusammen, um die Dynamiken des digitalen Zeitalters zu verstehen und an seiner Gestaltung mitzuwirken. Das „CAIS Kolleg“<sup>1</sup> fördert dabei innovative Projekte, die sich mit den gesellschaftlichen Chancen und Herausforderungen der digitalen Transformation befassen und Perspektiven für die Praxis entwickeln. In diesem Rahmen wird auch die hier beschriebene AG in den Jahren 2022 und 2023 gefördert. Ziel der AG ist es, didaktische Konzepte für die Nutzung digitaler Korpusinfrastrukturen in Lehre und Unterricht zu entwickeln und für verschiedene Vermittlungskontexte (Schule, Hochschule und außerschulische Fremdsprachendidaktik) adäquate Unterrichtsmodelle zu erarbeiten.

### Korpora als Bildungstechnologien: Möglichkeiten für den Unterricht und für die Lehrkräftebildung

Im Fokus der Entwicklungsarbeit stehen Konzepte des Forschenden Lernens, in denen Lernende an die aufgaben- und problembezogene Arbeit mit Korpora herangeführt werden. Dies beinhaltet die Gewinnung von Einsichten in die Systematik sprachlicher Strukturen sowie die Bedingungen und Ausprägungsformen von Sprachgebrauch und Sprachvariation. Die Lernenden sollen dadurch die Erarbeitung von Wissen und die Prüfung von Hypothesen zur deutschen Sprache als einen Prozess aktiver (Re-)Konstruktion kennenlernen, der sich auf die systematische Auswertung von Daten stützt, und ein Verständnis von Linguistik als empirisch arbeitende Wissenschaft entwickeln. Durch die Heranführung an den Umgang mit authentischen, durch Metadaten und Annotationen erschlossenen Sprachdaten sowie deren Auswertung anhand digitaler Recherche- und Analyse-schnittstellen entwickeln sie zudem ein Bewusstsein für die Anforderungen und Potenziale der Arbeit mit digitalen Forschungsressourcen zur deutschen Sprache (corpus literacy). Im

Bereich des Deutschunterrichts, des Unterrichts „Deutsch als Fremdsprache“ (DaF) sowie im Rahmen des Germanistikstudiums kann dies einen wichtigen, domänenspezifischen Baustein der Entwicklung von Daten- und Digitalkompetenzen (data literacy, digital literacy) bilden.

### CORPUS LITERACY UND DATA LITERACY ALS KOMPETENZBAUSTEINE FÜR EMPIRISCHE ZUGÄNGE ZUR DEUTSCHEN SPRACHE

Unter dieser Perspektive rücken Korpora, auch wenn sie vornehmlich als Ressourcen für die Forschung konzipiert sind und bereitgestellt werden, als Bildungstechnologien in den Blick. Durch den systematischen Einbezug in Unterrichtskontexte eröffnen sie der Untersuchung von Sprache, die als Kompetenzbereich in den Bildungsstandards für das Fach Deutsch nicht erst seit deren Überarbeitung 2022 verankert ist, eine Vielzahl an Möglichkeiten, Sprache in ihren Gebrauchskontexten und damit dort, wo sie als Instrument sozialer Verständigung authentisch greifbar wird, hinsichtlich ihrer Formen und Funktionspotenziale unter die Lupe zu nehmen. Die Arbeit mit Korpora stellt eine Ergänzung, nicht Alternative zur Arbeit mit Lehrwerken für den Unterricht dar: Sprachliches Wissen und Sprachbewusstheit, die durch die Arbeit mit Lehrwerken systematisch entwickelt werden sowie Annahmen zur Akzeptabilität, Domänen- und Textsortenspezifik sprachlicher Formen können anhand von Korpora an der Sprachwirklichkeit überprüft und über entdeckende Zugänge am empirischen Material bereichert werden. Dabei erfährt die Vielfalt und Flexibilität der Sprachverwendung in alltags- und beruflich relevanten Kontexten – in Texten, in Gesprächen und in der internetbasierten Kommunikation – eine Konkretisierung, die über Lehrwerke nur begrenzt hergestellt werden kann.

### ORIENTIERUNG AN KOMPETENZZIELEN DER BILDUNGSSTANDARDS

In den überarbeiteten Bildungsstandards für das Fach Deutsch in den Sekundarstufen (vgl. BiSta-D ESA-MSA 2022, S. 41) ist die Befähigung zur Arbeit mit (digitalen) Ressour-

cen zur deutschen Sprache als eigener Teilbereich des Kompetenzbereichs ‚Sprache und Sprachgebrauch untersuchen‘ verankert, die eine wichtige Grundlage für Kompetenzziele in anderen Bereichen bildet – etwa im Kompetenzbereich ‚Schreiben‘ für die Planung und Strukturierung von Texten (ebd., S. 24) oder bei der Untersuchung von Sprache für die Beschreibung sprachlicher Strukturen in Texten und Gesprächen, in gesprochener und geschriebener Sprache (ebd., S. 40). Das Ziel der Befähigung zur Arbeit mit (digitalen) Ressourcen zur deutschen Sprache korrespondiert dabei mit Teilzielen der von der Kulturministerkonferenz (KMK) formulierten Bildungsziele für die ‚digitale Welt‘, die unter anderem die Befähigung zur Recherche in digitalen Umgebungen (KMK 2017, S. 16) sowie zur Kenntnis und bedarfsgerechten Anwendung von digitalen Werkzeugen (ebd., S. 18) als verbindliche Kompetenzziele für den Unterricht festlegen (vgl. Beißwenger 2022).

Auch für den Bereich der Fremdsprachendidaktik (DaF) bildet unter anderem die Arbeit mit authentischem Sprachmaterial, das insbesondere auch Alltagssprachliche Kontexte einschließt, ein gewichtiges Argument für den unterrichtlichen Einsatz von Korpora (vgl. Edelhoff 1985; Imo/Weidner 2018; Gredel 2021).

### **Sprache datengestützt untersuchen: mündlich, schriftlich und digital**

Um die deutsche Sprache sprachsystematisch und gebrauchsbefugten zu untersuchen, verfügt die germanistische Linguistik über eine gut ausgebaute Landschaft von Korpora, die über digitale Ressourceninfrastrukturen nachhaltig bereitgestellt werden und die über WWW-basierte Nutzerschnittstellen recherchiert und abgefragt werden können. Diese bieten eine Vielzahl an Möglichkeiten, schulisch relevante Analyseebenen und Gebrauchskontexte von Sprache zu untersuchen: Lexik, Semantik, Wortbildung, Wortgrammatik, Syntax, Orthographie, Pragmatik, sprachliche Variation (synchron und diachron), Unterschiede von mündlicher und schriftlicher Sprachverwendung, von Gesprächen, Texten und internetbasierter Kommunikation.



Die Mitglieder der Arbeitsgruppe. Vordere Reihe v.l.n.r.: Dr. Ziko van Dijk (Wikimedia / Klexikon), Prof. Dr. Carolina Flinz (Università degli Studi di Milano), Nadja Radtke, M. A. (Lichtenbergschule Darmstadt), Prof. Dr. Eva Gredel (Universität Duisburg-Essen), Dr. Lothar Lemnitzer (Berlin-Brandenburgische Akademie der Wissenschaften); hintere Reihe v.l.n.r.: Prof. Dr. Michael Beißwenger (Universität Duisburg-Essen), Dr. Harald Längen (IDS), Dr. Thomas Schmidt (LinguisticBits), Lena Josianne Rebhan, M. A. (Universität Duisburg-Essen), Dr. Thomas Bartz (Universität Duisburg-Essen), Dr. Jan Oliver Rüdiger (IDS); nicht im Bild: Antonia Hamdi, B. A. (Universität Duisburg-Essen), Sarah Steinsiek, M. A. (Universität Duisburg-Essen), Prof. Dr. Angelika Storrer (Universität Mannheim)

## **IM FOKUS DER AG: SPRACHKORPORA DES IDS UND ANDERER ANBIETER**

Für die Entwicklung von Materialien und didaktischen Konzepten für den Deutschunterricht und für die Lehrkräftebildung Deutsch/DaF nutzt die AG unter anderem die folgenden Korpusressourcen:

- das **Deutsche Referenzkorpus (DEREKO)**, das am Leibniz-Institut für Deutsche Sprache (IDS Mannheim) u. a. über das Corpus Search, Management and Analysis System (COSMAS II) als Abfrage- und Rechercheschnittstelle zugänglich ist und das mit über 55 Milliarden Wörtern (Stand: 3.5.2023) die weltweit größte linguistisch aufbereitete Sammlung digitaler Korpora zum geschriebenen Gegenwartsdeutschen darstellt, die eine große Bandbreite an Textsorten und Domänen schriftlichen Sprachgebrauchs abdeckt. Neben redigierten Texten umfasst DEREKO dabei auch Korpora zu Diskussionsseiten in der deutschen Wikipedia sowie zu weiteren Formen internetbasierter Kommunikation, in denen geschriebene Sprache für die Realisierung sequenziell organisierter Interaktion eingesetzt wird. Dies ermöglicht es, mit Lernenden Muster und Praktiken in digitaler, ‚interaktionsorientierter‘ Schriftlichkeit (Storrer 2018) zu analysieren und diese mit der Sprachverwendung in Texten zu vergleichen. Beispiele für die Nutzung dieses Korpus für wortbezogene Untersuchungen aus dem Kreis der AG-Mitglieder, von denen ausgehend Didaktisierungen für den Unterricht denkbar sind, sind Gredel (2018) und Gredel/Flinz (2020) zu Neologismen (*Newstickeritis*, *Löschomanie*) sowie Storrer/Herzberg (2022) zu *okay*. <<https://cosmas2.ids-mannheim.de/cosmas2-web/>>.

- die Korpusammlung im Portal „Der deutsche Wortschatz von 1600 bis heute“ des Projekts „**Digitales Wörterbuch der deutschen Sprache**“ (DWDS) am Zentrum Sprache der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW), die Korpora zur geschriebenen deutschen Sprache in Vergangenheit und Gegenwart – vom 17. Jahrhundert bis zur Sprache in sozialen Medien (z. B. Blogs) – umfasst. Die DWDS-Korpora ermöglichen damit eine auch diachrone Untersuchung der deutschen Sprache. Durch die Vernetzung mit digitalen Wörterbüchern zum Deutschen und mit innovativen Werkzeugen für die Auswertung und Visualisierung von Korpusbelegen im diachronen Verlauf bietet die DWDS-Plattform spannende Möglichkeiten u. a. für die Analyse von Fragen des Wortschatzes und des lexikalisch-semantischen Wandels. Ein Beispiel für die Untersuchung ausgewählter Wortschatzeinheiten im Rahmen des didaktischen Konzepts einer Wörterbuchwerkstatt aus dem Kreis der AG-Beteiligten beschreibt Bartz (2016). <[www.dwds.de](http://www.dwds.de)>.
- das **Forschungs- und Lehrkorpus Gesprochenes Deutsch (FOLK)**, das über die Datenbank Gesprochenes Deutsch (DGD) nutzbar ist und das Daten gesprochener Sprache (Audioaufnahmen und damit alignierte sprachanalytische Transkripte) aus unterschiedlichen Sprachgebrauchsdomänen (Beruf, Familie, Unterricht) umfasst. Eine Nutzerstudie hat bereits gezeigt, dass FOLK etwa für die Sprachvermittlung im Bereich DaF/DaZ genutzt wird (vgl. Fandrych et al. 2016). Aufgrund der Beschränkung auf wissenschaftliche Nutzungskontexte kann FOLK zwar nicht im Schulunterricht eingesetzt werden; für die Lehramtsausbildung bietet die Resource allerdings innovative Möglichkeiten, die Organisation und sprachliche Gestaltung mündlicher Interaktionen sowie syntaktische und lexikalische Besonderheiten der gesprochenen Sprache zu untersuchen und die fachliche Kompetenz von Lehramtsstudierenden in Bezug auf die Spezifik mündlicher vs. schriftlicher sowie interaktionaler vs. monologischer Formen der Sprachverwendung weiterzuentwickeln (vgl. das Kompetenzfeld „Mündliche und schriftliche Kommunikation“ im Kompetenzbereich „Sprache und Sprachgebrauch untersuchen“ der Bildungsstandards Deutsch/BiSta-DESA-MSA 2022, S. 38-39). <[https://dgd.ids-mannheim.de/dgd/pragdb.dgd\\_extern.welcome](https://dgd.ids-mannheim.de/dgd/pragdb.dgd_extern.welcome)>.
- die **Mobile Communication Database (MoCoDa 2)**, die eine kontinuierlich erweiterte Sammlung von WhatsApp-Chats enthält, anhand derer sich über eine spezifisch auf die Besonderheiten dieses Datentyps zugeschnittene Such-, Filter- und Abfragemöglichkeiten Merkmale der Sprachverwendung in digitaler Alltagsschriftlichkeit analysieren lassen. Beispiele für die Nutzung der MoCoDa2 für Untersuchungen zu Spezifika und Besonderheiten des interaktionsorientierten Schreibens aus dem Kreis der AG-Mitglieder, die Ausgangspunkte für die Entwicklung von Unterrichtsmodellen bilden können, sind die Arbeiten zu sprachlichen und pragmatischen Besonderheiten digitaler Messaging-Kommunikation (Beißwenger 2018, Beißwenger/Pappert 2020) sowie – unter Einbeziehung u. a. auch des FOLK-Korpus für die gesprochene Sprache – Steinsiek (2021) zur Interjektion *HM* und zu Verwendungen von *weil* mit Verbzweitstellung. <<https://db.mocoda2.de/>>.
- das **Deutsche Textarchiv (DTA)**, das einen disziplinen- und gattungsübergreifenden Grundbestand deutschsprachiger Texte mit einem Schwerpunkt auf dem frühen 17. bis zum frühen 20. Jahrhundert bereitstellt. Entsprechend seiner Konzeption und Zusammenstellung ermöglicht das DTA systematische, textgestützte Zugänge zur Sprach- und Textsortengeschichte des neueren Deutschen und kann in Unterrichtskontexten zum Beispiel für Einblicke in sprachgeschichtliche Entwicklungen genutzt werden, anhand derer sich Tendenzen des Gegenwartsdeutschen auch in historischer Dimension betrachten und erklären lassen. <[www.deutsches-textarchiv.de](http://www.deutsches-textarchiv.de)>.

## Dokumentation der Ergebnisse

Die Ergebnisse der AG werden durch zwei Publikationen dokumentiert, die 2023 bzw. 2024 erscheinen sollen: eine Themenausgabe der Zeitschrift KorDaF (Korpora Deutsch als Fremdsprache) zum Thema „Korpora internetbasierter Kommunikation im DaF-/DaZ-Unterricht“ sowie eine Buchveröffentlichung, die sich an Lehramtsstudierende und Lehrkräfte für das Unterrichtsfach Deutsch sowie an Hochschullehrende in Lehramtsstudiengängen mit Deutsch/Germanistik richtet:

- Trotz der hohen Relevanz digitaler Kommunikation und der Zugänglichkeit von Korpora findet der Einsatz von Korpora internetbasierter Kommunikation (IBK-Korpora) im DaF-/DaZ-Unterricht bislang nur vereinzelt statt und wurde auch noch nicht systematisch betrachtet. Die geplante Themenausgabe der Zeitschrift KorDaF bearbeitet dieses Desiderat an der Schnittstelle von korpuslinguistischer Forschung und dem Transfer in das Praxisfeld des DaF-Unterrichts, indem sie nach Eigenschaften von IBK-Korpora fragt, die diese für den Einsatz in der Fremdsprachendidaktik qualifizieren. Dabei sollen nicht nur konzeptionelle Überlegungen, sondern auch ganz konkret didaktische Szenarien und Unterrichtsmodelle für die Nutzung von IBK-Korpora im Unterricht im Fokus stehen. Die Veröffentlichung der Ausgabe ist für Ende 2023 geplant.

## PRAXISRELEVANTES WISSEN FÜR LEHRKRÄFTE: ZWEI PUBLIKATIONEN IN VORBEREITUNG

- Auch im Hinblick auf die Qualifizierung von Lehrkräften für das Unterrichten mit digitalen Technologien und Ressourcen – im Bereich der Deutschdidaktik und des Deutschunterrichts insbesondere für die Arbeit mit digitalen Sprachressourcen im Unterricht – besteht großer Bedarf an der Vermittlung von Basiskompetenzen und unterrichtspraktischen Hinweisen für die didakti-

sche Nutzung von Korpora. Die Monographie „Sprachkorpora im Deutschunterricht“ soll im Stil eines Einführungsbuchs, das in Veranstaltungen der Lehramtsstudiengänge oder auch zum Selbststudium verwendet werden kann, in die fachlichen (linguistischen, korpuslinguistischen, sprachdidaktischen) Grundlagen für den Einsatz von Korpora im (Deutsch-)Unterricht einführen und Lehramtsstudierenden und Lehrkräften exemplarisch Unterrichtsmodelle für die korpusgestützte Untersuchung von Sprache und Sprachgebrauch an die Hand geben, in denen die beschriebenen Korpusressourcen zum Einsatz kommen. Das Erscheinen des Bandes ist für das Jahr 2024 geplant. ■

## Anmerkung

- <sup>1</sup> Das CAIS Kolleg wird auf der Internetpräsenz des CAIS detaillierter beschrieben. <[www.cais-research.de/cais-kolleg/](http://www.cais-research.de/cais-kolleg/)>.

## Literatur und Quellen

- Bartz, Thomas (2016): Digitale Sprachressourcen im Deutschunterricht: Korpus-basierte Recherche und Analyse in der ‚Wörterbuchwerkstatt‘. In: Chudak, Sebastian/ Drumbl, Hans/Nardi, Antonella/ Zanin, Renata (Hg.): IDT 2013. Band 6 – Sektionen F2, F3, F4. Medien in Kommunikation und Unterricht. Bozen: Bozen-Bolzano University Press, S. 237-248.
- Beißwenger, Michael (2018): WhatsApp, Facebook, Instagram & Co.: Schriftliche Kommunikation im Netz als Thema in der Sekundarstufe. In: Gailberger, Steffen/Wietzke, Frauke (Hg.): Deutschunterricht in einer digitalen Gesellschaft. Unterrichtsarrangements für die Sekundarstufen. Weinheim: Beltz Juventa, S. 91-124.
- Beißwenger, Michael/Pappert, Steffen (2020): Sprachverfall durch Emojis? Eine pragmalinguistische Perspektive auf den Beitrag von Bildzeichen zur digitalen Kommunikationskultur. In: Aptum. Zeitschrift für Sprachkritik und Sprachkultur 16, 1, S. 32-50.
- [BiSta-D ESA-MSA 2022] Sekretariat der Kultusministerkonferenz (Hg.) (2022): Bildungsstandards für das Fach Deutsch. Erster Schulabschluss (ESA) und Mittlerer Schulabschluss (MSA). <[www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2022/2022\\_06\\_23-Bista-ESA-MSA-Deutsch.pdf](http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2022/2022_06_23-Bista-ESA-MSA-Deutsch.pdf)> (Stand: 4.12.2022).

- Edelhoff, Christoph (1985): Authentizität im Fremdsprachenunterricht. In: Edelhoff, Christoph (Hg.): *Authentische Texte im Deutschunterricht. Einführung und Unterrichtsmodelle*. München: Hueber, S. 7-30.
- Fandrych, Christian/Frick, Elena/Hedeland, Hanna/Iliash, Anna/Jettka, Daniel/Meißner, Cordula/Schmidt, Thomas/Wallner, Franziska/Weigert, Kathrin/Westpfahl, Swantje (2016): User, who art thou? User profiling for oral corpus platforms. In: Calzolari, Nicoletta/Choukri, Khalid/Declerck, Thierry/Goggi, Sara/Grobelnik, Marko/Maegaard, Bente/Mariani, Joseph/Mazo, Hélène/Moreno, Asunción/Odijk, Jan/Piperidis, Stelios (Hg.): *Proceedings of the Tenth International Conference on Language Resources (LREC'16)*, Portorož, Slovenia, S. 280-287.
- Gredel, Eva (2021): CMC-Korpora und Digital Literacy in der Fremdsprachendidaktik: Relevanz, Potentiale und didaktische Szenarien. In: *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 26, 1, S. 109-135.
- Gredel, Eva (2018): *Itis*-Kombinatorik auf den Diskussionsseiten der Wikipedia: Ein Wortbildungsmuster zur diskursiven Normierung in der kollaborativen Wissenskonstruktion. In: *Zeitschrift für Angewandte Linguistik* 68, 1, S. 35-72.
- Gredel, Eva/Flinz, Carolina (2020): Morphosyntax im deutsch-italienischen Vergleich: Eine kontrastive Fallstudie zu Wortbildungsprodukten und deren diskursiven Funktionen in der internetbasierten Kommunikation der Online-Enzyklopädie Wikipedia. In: *Deutsche Sprache* 48, 3, S. 193-209.
- Imo, Wolfgang/Weidner, Beate (2018): Mündliche Korpora im DaF- und DaZ-Unterricht. In: Kupietz/Schmidt (Hg.), S. 231-251.
- [KMK 2017] Sekretariat der Kultusministerkonferenz (Hg.) (2017): Strategie der Kultusministerkonferenz „Bildung in der digitalen Welt“. Beschluss der Kultusministerkonferenz vom 8.12.2016 in der Fassung vom 7.12.2017. <[www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/2018/Digitalstrategie\\_2017\\_mit\\_Weiterbildung.pdf](http://www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/2018/Digitalstrategie_2017_mit_Weiterbildung.pdf)> (Stand: 4.12.2022).
- Kupietz, Marc/Schmidt, Thomas (Hg.): *Korpuslinguistik. (=Germanistische Sprachwissenschaft um 2020, 5)*. Berlin/Boston: de Gruyter.
- Steinsiek, Sarah (2021): „Hm aber glaube das das Mega schwierig ist weil es soll ja auch ein bisschen schwierig sein.“ Korpusgestützte Untersuchungen zu grammatischen Besonderheiten in digitaler Alltagskommunikation. MA-Arbeit, Universität Duisburg-Essen. <<https://doi.org/10.17185/dupublico/75213>>.
- Storrer, Angelika (2018): Interaktionsorientiertes Schreiben im Internet. In: Deppermann, Arnulf/Reineke, Silke (Hg.): *Sprache im kommunikativen, interaktiven und kulturellen Kontext*. Berlin/Boston: de Gruyter, S. 219-244.
- Storrer, Angelika/Herzberg, Laura (2022): Alles okay! Korpusgestützte Untersuchungen zum Internationalismus OKAY. In: Beißwenger, Michael/Lemnitzer, Lothar/Müller-Spitzer, Carolin (Hg.): *Forschen in der Linguistik. Eine Methodeneinführung für das Germanistik-Studium*. Paderborn: Brill/Fink, S. 37-59.

### Bildnachweis

Esther Laukötter (CAIS NRW). ■